ÇANKAYA UNIVERSITY FACULTY OF ENGINEERING COMPUTER ENGINEERING DEPARTMENT

Project Report

Version 1

CENG 408

Innovative System Design and Development II

P201919

Image Analysis for the Classification of Brain Tumor Location on MR Images

Nisan Çandır -201611010

Ece Yılmaz -201611064

Kevser Sürgüveç- 201411055

Deniz Yıldız -201511066

Advisor: Faris Serdar TAŞEL

Co-Advisor: Roya CHOUPANI

Abstract

In today's world, after heart and lung diseases, cancer is the second biggest health issue that causes death. According to World Health Organization data; because of various cancer types particularly lung cancer; only in 2018, 9.6 million people lost their lives. In recent technological surveys that developed with latest technological tools, it is ascertained that detecting cancer and tumor on early stages has important value on extending patients lives. MR imaging is one of the technological tool used in medical that helps doctors to detect and see healthy and sick tissues. It has several advantages over other imaging techniques, providing high contrast between soft tissues. There has been various algorithms developed to process MR images. These algorithms gave different results depending on datasets and anatomic structures of images. Increasing computation of these algorithms and getting fattest, accurate results will make the patients start early treatment and will get ahead of possible bad outcomes. In this study, we aimed to use developed algorithms to detect brain tumor area to develop a new system. This way we can create a system to help doctors to detect brain tumors in early ages. We did large researches on image processing algorithms that used on MRI ımages to detect brain tumor regions. After

the research, it shows that Neural Network with backpropagation or SVM with Hybrid method gives best results. Neural Networks need large datasets and more complicated. Since, we don't have any experiences on image processing, we decided to use and implement Hybrid Method with SVM. We will use TCIA's, kaggle's and BraTS's websites to find the datasets to use for this project. After doing a large research, we learned how most of the image processing algorithms that used for detecting brain tumor region and brain tumor type works. Such as K-means, Fuzzy, SVM, Neural Networks, K-NN, Region Growing.) We also learned that Python is best way to implement these methods and we gained a knowledge about how to implement these methods.

Key words:

Brain Tumor, MRI Images, Machine Learning, Image Processing, Artificial Intelligence, Deep Learning, Segmentation, Classification

1.Introduction

1.1 Problem Statement

Image processing is important subject used on medical innovations. Detecting brain tumor region is a critical and sensitive topic, because patients have more chance getting better or getting completely cured if they start early treatment. Therefore, early detection of tumors is very important. Mostly, image processing methods are used for detecting tumor region. There has been many experiments on detecting tumor region with image processing algorithms in past years. Even though, is has been found that some algorithms and methods works better than other. We did a broad research on this topic and we found out that, there hasn't been found one image processing algorithms that outperforms other algorithms. We are interested in image processing field and we wanted to improve ourselves in that area. So we thought working on these experiments would be better for us beginners.

1.2 Background or Related Work

Mostly, other researchers and students, professors who are interested in this field, made experiments on these subject and released articles in past years. The image processing methods can be used on and related to many other fields such as medicine, artificial intelligence, computer vision. Therefore, there is a lot of researches, articles and project related to this field. Our project is mostly related on image processing in medicine. Future experiments on this field could use our project and our articles as reference to our project and use our project to develop it. Also doctors who wants to get a second judgement on their diagnose can use our system.

1.3 Solution Statement

After doing research, we found out that Neural Networks or SVM with Hybrid Method gives the best result. We decided to use SVM with Hybrid method on this project. For Hybrid method, we decided to use K-means and Fuzzy algorithm. We decided to take some of the articles we found to consideration and implement them as reference. We will implement the methods used in these articles and try to improve them if possible. We will implement these methods on python programming language

1.4 Contribution

We will combine some of the articles we found on our research. After combining the articles. The algorithms used on articles will work better and will give better results. Since, it is proven that Hybrid Method works better with past experiments. We also plan to develop these methods based on accuracy and computation time.

2. Literature Search

2.1 Examples of Segmentation Algorithms

Thresholding: Thresholding has been one of the images segmentation strategies which is widely utilized. This approach is acceptable for images of growing intensity levels. The image is divided specifically into various regions to use this form, depending on the intensity values [11]. Segmentation based on edges : Edge segmentation approaches segment an image depending on extreme differences between close-edge intensity values [29]. The outcome is an image in binary form[11].

Region Growing Technique: Region Growing Technique is among Its most regular common form of splitting. The technique begins with such a kernel pixel and raises the area by inserting threshold value-based adjacent pixel. Whenever an area's growth slows down, another seeds pixels is picked that does not relate to any area, as well as the cycle is repeated [14].Increasing area is halted if all pixel contribute to a certain area. For specific region- widening segmentation can be used for delineating thin basic structures like tumors and lesions[15][11].Reaches for area-growing leverage the interesting point is that close-up pixel have identical grey value. Begin with such a pixel (kernel) and instead slowly add different pixels

1. Pick kernel pixel
2. Test and attach any adjacent pixels to that of the area if they're close to the kernel
3. Do step 2 for every one of the updated pixels; quit if you can't add any more pixels[13].

The specific drawbacks to use that approach are-Manual intervention is often needed to pick the seeds level. JAILBREAK-Noise-sensitive for generating gaps and over-segmentation of isolated areas [16][11].

-The key conclusion of such a method that its picture amplitude is almost unchanged throughout areas.

-To anti-smoothly changing areas (e.g. layered areas), that method would not operate.

-Further advanced methods are needed to enable for further practical differences in brightness attributable to shading[13].

Genetic Algorithm: (GA) is a method that an inspiration for biological evolution. Genetic Algorithms rely on bio-originated operators. These operators define optimization function but also does crossover,mutation and selection etc. [36].

Classifiers:

* KNN K-Nearest Neighbor (k-NN): k-NN algorithm is one of simplest method that gives reliable result on accuracy [17][18]. K-NN has slow run time but also gives better accuracy and reliability on MR Images than most of the other classify methods

The phases of this method is

-Step1:Specify the value of k

-Step2: Calculate the differences of all the instruction samples and the question instance

-Step3: Filter the range by the minimum distance kth

-Step4: Put the class of plurality to

-Step5: Decide that category

-Step6: Segment the abnormalities in the brain [11]

SVM (Support Vector Machine): SVM minimizes the limitations of generalization error instead ıf minimizing objective function of training examples. To conclude, when SVM is used on data that is not from training data set, it gives good result. This results in a , an SVM tends to perform well when modified to data outside of a prepare to practice.SVM retains this benefit by concentrating on the most difficult to identify training cases. Some examples of "borderline" preparation are dropping support vectors. SVM is based on the hyperplane. Hyperlanes tries to maximize the differences of the two classes.

SVM takes data as input and determines whether the outputs belongs to one of two class and does this for each given input.This is called realistic classification. SVMs can also efficiency an effective nonlinear classification usage of kernel trick. To maximize the margin between categories, the empirical risk and classification is managed using its kernel There have been several kernel functions including linear, polynomial degree, and the function Radial Basis (RBF). [20]. SVM has two phases; testing and training. In training stage, SVM method trained by properties from its learning algorithm is given as an input During learning SVM chooses the correct margins between two groups . Characteristics are labelled with particular group within associative category . Artificial neural network Just a lot problems with local minima and number of selected neurons for each question. To solve this problem, SVM does not occupy Minimum locality, and neuron selection upload by introducing the concept of hyper flugzeuge [11]. Clustering A clustering was one of those MRI's most effective segmentation techniques, where pixels are clustered into groups without prior information or preparation. That classifies pixels of the strongest Possibilities within the same group. Education in clustering method is done using pixel characteristics with properties all class [21][11].

•K-Means: K-means clustering algorithm is the easiest unattended genetic algorithm capable of solving the clustering issue. The method pursued is very easy for classifying a defined array of results across a variety of groups.Centers are identified in K-means ' K, ' one for each cluster. Such groups requires to be located far from everything else . The next step is to take a point that belongs to a set of data and align it with the closest core. The first phase is completed when no stage is pending, and fast restructuring is finished. The next phase is to recheck ' k ' new centres as the resulting barycenter of the groups from the beforehand step. Since making ' K ' new centroids, there must be a new connection between the same data setpoints and the nearest new centre. A loop was formed. As a result of this process, the k centers are gradually changing their position until centers are no longer moving.

Fuzzy C-Means: FCM bundling is an unmonitored techniques for the examines of data. This method applies members for each data set referring to each group core,based on the range between both the group core and the piece of data.The data point near the cluster center is more connected to the particular node. In general, the membership description for all will be equivalent to the stage [11].FCM algorithm gives good results for overlapping data as well as runs better than the method K means The piece of data may be a part of a group core there. FCM's biggest drawback is: In all clusters, the overview for the member price of a point xi would be one but the external parts are worth further participation. Therefore, the method has issues with anomaly points. The group core tendentially go to the center of all dataset because of the presence of all group members[22]. This considers only the strength of the image while generating unfriendly outputs in bruising images. A blend of approaches of methods are provided to make FCM stable and homogeneous contra noise, but it is not yet good.

Adaptive Fuzzy Clustering: A modified version of regular FCM is the adaptive fuzzy clustering algorithm[22]. In this process the membership values are determined using equation:

$$\mu_{ij} = \frac{n * \left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{p}\sum_{z=1}^{n}\left(\frac{1}{d_{kz}}\right)^{\frac{1}{m-1}}}$$

This algorithm is efficient in handling outlier points of data. This gives very low membership for outlier points compared with the FCM algorithm[22].

Artificial Neural Network: Segmentation based on Neural Network is completely various from traditional segmentation methods. In this, an image is transformed throughout the a Neural Network at first. Where each Neuron states for a pixel[24], the image segmentation problem is thus translated into a issues of power minimisation.

With training sample collection, the neural network was trained to specify the relation and powers all nodes Segmentation of the neural network involves Two decisive moves: take-out and image segmentation building on the neural network. Feature removal is so much important because it decides the data entry of the neural network[25], first of the photos remove most of the apps so this is they are appropriate to section and then the neural network entry . All of the features chosen are composed of cluster boundary particularly un- linear function domain. Neural segmentation simple on a network has three basic characteristics: Strong parallel efficiency and swift computing[24]. Improve the effects of segmentation when the data deviates from standard situation[24]. Reduced need for expert input during process of image segmentation. However there are other drawbacks to segmentation based on neural networks, as for example:

Any type of details regarding segmentation need to be identified in advance. Neural network need to be educated in advance making use of mental workout [24]. Training cycles can be very long, and at the same time, overtraining should be avoided.

Artificial Neural Network can be split onto two factions according to the architecture: Feed-Forward Network Feed-Backward Network or Recurrent Network The neurons are organized in layers in feed-forward network these have bidirectional ties and between two. Feed forward networks only generate only number of calculated variables for output. It is known as a stationary system, since beliefs of the output using current input only. Sometimes called memoryless network power the Network onwards .The nerves are organized in shallows in feedback network which have bidirectional Relationships between them. Returns
networks generate a set of werts that are modified simple on the fed outputs. Often referred to as Fluid network is the feedback network, because the result values addiction on the Past Statement values[11].

Back Propagation Algorithm: Back Propagation algorithm [26] is available in feedforward swallowed ANN. This method operates with continuous performance for the feed-ahead networks. The nerves come in waves n this network and transmit their Indicates in the Way toward. That caused mistakes were peddled backwards. The network generates feedback from cells in the entry surface, and the network's feedback on a feedback layer is given by the nerves. The Network is composed of one or more intermediate hidden layers. Requires supervised instruction in Back Propagation Algorithm, i.e. the algorithm is given examples of both the entry and the output to be measured. Failure is calculated between input and estimated output. The network is equipped Values are natural and sizes are changed later in order to get the minimum error. If the error is minimal the network will be perfect. Through back propagation, each time an example is introduced the weights and thresholds are modified, so that the error slowly reduces.That is repeated until the error [11] does not shift. The human brains are built on feedback, The segmentation of MR images relies very heavily on the learning collection used in the their controlled instruction. The learning set is built by choosing feature columns from a specific MR image or an aggregate of MR images and reflects the decision of the actual expert who allocate tags to the attribute vectors as per the substances they reflect [11][27].

Hybrid Techniques: Similar to the specific procedure, imaging modalities, body area and other variables, methods for conducting segmentations vary greatly. There is currently no specific method of segmentation which can yield acceptable results for every medical image. Combining several segmentation techniques together to create a hybrid system can sometimes dramatically improve the performance and robustness of the segmentation compared to each individual unit.The hybrid segmentation approach integrates boundary and regional segmentation methods which amplify the strength but reduce the weakness of both approaches.[28].

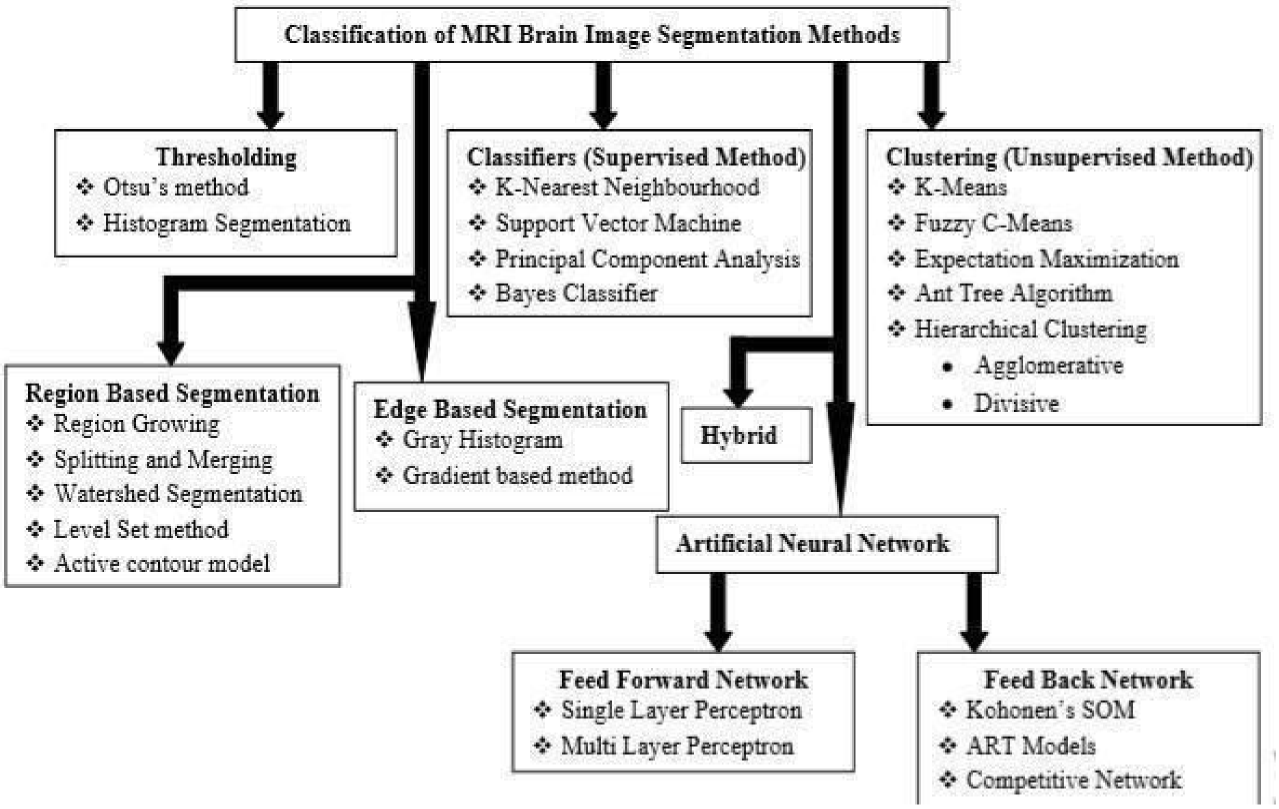

Figure 1: *Diagram of Image Processing methods used for MRI Brain Tumor Detection*

2.2 Related Literature

S. Huda et al. [39] proposed a worldwide integrated designed significant Variety method for tumor function that is merged with Tackling and Decision Taking tree to identify oligodendrogliomas with the chemosensitive attribute and to reduce the effect of non-homogeneous or imbalanced dataset. Their algorithm as called GANNIGMA (Globally Optimized Artificial Neural Network Input Gain Measurement Approximation) hybrid feature screening designed to find out important properties of image that developed to create a more simple rule. The proposed hybrid model incorporated information about the partnership between the solutions to filters and wrappers. This approach helped to provide a simplified decision rule and to obtain higher accuracy. Hybrid feature selection gives the highest accuracy results. However, it has the disadvantage of complexity, thus this may lead to difficulties at implementation steps. If there may occur a problem at some step of processing, stepping backward also causes wasting of space to go.

The goal of the proposed analysis is continuing an automated algorithm for the segmentation of brain tumors simple on the Fuzzy C-means group and glowworm swarm optimization (GSOFCM) and region growing technique. The proposed approach has of these stages:

-Step-1 is accelerating FCM clustering for tissue segmentation process based on GSO.

-Step-2 is an abnormal detection process that helps to check the results of GSOFCM method by fuzzy symmetric measure (FSM).

-Step-3 is that segmentation the tumor region from abnormal slices by region growing technique. GSO for selecting the initial centroids for the FCM method and segments the brain certain muscles as background, gray matter (GM), white matter (WM), cerebrospinal fluid (CSF). Then abnormal tracking process is done by FSM.

Finally, the region growing method is utilized in abnormal slices for segmenting the tumor region.

The findings show that the approach suggested provides adequate findings for segmentation of the brain tumors.The computation time of tumor segmentation algorithms ranged from 1 to 2 minutes that depends on the size of the dataset. The study results that are likened with the other forms is in the table below: [33]

The proposed method resulted better at accuracy compared to other methods except [34] method and has faster computation time.

Not only can the proposed method identify the total tumor area but can also precisely Structure section of the intratumor [35]. The theoretical research was focused on a flowed machine learning convolutional neural network composed of two subnetworks: (1) a Tumor Localization Network (TLN) and (2) an ITCN(IntraTumor Convolutional Network) intratumor classification network. The first subnetwork aim was to determine the tumor area from a slice of an MRI. FCN-8s(Fully- Convolutional Network) is used for TLN. The ITCN was then utilized to mark the tumor area identified into assorted subarea. In particular, ITCN has taken advantage of a convolutional neural network (CNN) with deeper architecture and smaller kernels. The proposed work has been Authorized on datasets of multimodal brain tumor segmentation (BRATS 2015), containing 220 cases of high grade glioma (HGG) and 54 cases of low grade glioma (LGG). Dice similarity coefficient (DSC), positive predictive value (PPV), and sensitivity were used as evaluation metrics. Their findings revealed that their approach could produce positive segmentation outputs and have a quick segmentation rate. More precisely, on the combined (HGG + LGG) test set, the proposed approach won comparable and overall better DSC values (0.89, 0.77, and 0.80) compared with other methods [35].

All Magnetic Resonance Images (MRI) were normalized to 64x64 matrix of 10 mm slice thickness yielding a homogeneous element of 16 using Discrete Wavelet Transform (DWT) [40]. If demonstrates tumor, the neighboring 3 slices are analyzed to locate tumor margins and hence verify the prediction. If slices 6-7 and 11-12 were qualify is above , then all slices from 6-12 are considered segmentation.

The left and the right sides of the brain are examined to determine if there is a tumor. They tested the algorithm with Principal Component Analysis (PCA) and without Principal Component Analysis (PCA), and found that there was no significant change in accuracy whereas the computation time increased increased significantly with its use. Magnetic Resonance Imaging that is uploaded from medical image archives are former .mha format files [41]. MATLAB is not deal with .mha format files. Then it was converted to .jpeg format. Applied MR image Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) technique for the extraction of features. After that, kernel Support Vector Machine classifier was built which is required for classifying images as normal or abnormal. This approach was a hybrid of DWT (Discrete Wavelet Transform) Used only for removal function, then the principal component analysis (PCA) for decrease the function and for the classification of MR images the Support Vector Machine has been used.

They remove the background of FLAIR image using the T1-weighted map [42]. For initial segmentation, Fuzzy C-Means (FCM) algorithm was used on intensity features. Fuzzy C- Means (FCM) algorithm segmented into pre-specified number of clusters.

Fuzzy C-Means (FCM) described the degree of similarity of one pixel each cluster. They have compared the tumor region with T1-weighted and removed the mismatch regions. If slice did not contain any infarct region, they stopped the algorithm and declare slice as normal. This algorithm did not assume the shape and size of the tumor regions. All local and global knowledge was integrated into this system. This method incorporates additional information present in T1-weighted which improves performance in the presence of artifacts and helps boundaries to be improved.

In this proposed approach [43], they have utilized MR images of the brain, segmented brain tissues into normal tissues such as white matter, gray matter, cerebrospinal fluid, and tumor-infected tissues. They Preprocessing Included algorithms to improve the signal-to-noise ratio and to eliminate the effect of undesirable noise effects. Furthermore, they used Berkeley wavelet transform. They compared 4 different methods. These methods K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), ANFIS, Back Propagation. They found Support Vector Machine (SVM) algorithms accuracy highest than K Nearest Neighbors (K-NN), ANFIS and Back Propagation. In this paper, they used 4 different methods. In the result of experiments, SVM has the highest rate of specify, sensitivity and accuracy rate. K-NN and Back Propagation has the lowest specificity rate.

This work [44] aimed at demonstrating that consideration of the anatomical structure for normal-abnormal classification will improve classification outcome. Existing work has shown that statistical features of the feature vector, gray level co-occurrence matrix (GLCM) alongside Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) have yielded better results than other methods.

Segmentation have been done by using Multi Fractal. Then classification methods was applied using two techniques that were Adaboost Classifier and machine learning based Back Propagation Neural Network. In Adaboost classifier, the tumor was classified using Intensity Features whereas in BPNN, classification was done based on the tumor region. MLBPNN has an additional advantage compared to Adaboost in determining whether the tumor is in early or advanced stage. During this research method, the Back Propagation Neural Network based on machine learning uses 30 sample images to make the classification of tumor type.

Accuracy:

Adaboost: 67.15 / MLBPNN: 93.3

Sensitivity: Adaboost: 68.18 / MLBPNN: 71.42

Specificity: Adaboost: 53.04 / MLBPNN: 88.88

In this work [45], they proposed a model with a template-based K-means and an improved Fuzzy C-means (TKFCM) algorithm for detection of brain tumors. In this proposed approach, firstly, the template-based K-means algorithm was applied to significantly initialize segmentation by selecting the template perfectly, based on gray-level image intensity. After that, the modified membership has been determined by the distances from cluster centroid to cluster data points using the Fuzzy C-means(FCM) algorithm while contacting its best result, and the improved FCM clustering algorithm was used to detect tumor location by updating the membership function that was obtained based on the different tumor image features including Contrast, Energy, Dissimilarity, Homogeneity, Entropy, and Correlation.

Simulation results demonstrated that under limited detachment of gray-level strength, the proposed algorithm achieved better detection of abnormal and normal tissues in humans brain.
When they compared TKFCM with other algorithms, Sensitivity:
TKFCM 97.43% | Thresholding 76.9% | ANN9 5.5% Computation time: Thresholding 3 min | TKFCM 40-50 s TKFCM know 38 abnormal image correctly in 40 data image.
Improved Fuzzy C-Means algorithm gave better results with template based K-Means algorithm when they have compared them with other algorithms.
This work [46] described the method used to segment MRI image of brain use of Hierarchical Self-Organizing Map (HSOM). In proposed work, they improved segmentation method consisting of two phases. In the first phase, the MRI brain image has been acquired from patients' database. In that film, artifact and noise were removed.

After that, HSOM was applied for image segmentation. The HSOM was the extension of the conventional self organizing map that is used to classify the image row by row. In this lowest level of weight vector, a higher value of tumor pixels, computation speed has been achieved by the HSOM with vector quantization. They measured the amount of cancer cells in various pixels in neighborhood 3x3, 5x 5, 7x7, 9x9, 11x11 windows. Execution time of them was 13.76 in 795 number of segmented pixel when winning neuron 209, which made this tool has been better than other segmentation tools. The focus area was divided and the assessment of this method by the doctor with whom the plan was non-operated was optimistic and this method supported the doctors in the assessment, the planning of the care plan and the tumor monitoring process.

Twenty state-of - the-art tumor segmentation methods were modified to a collection of 65 low- and high-grade glioma multi-contrast MRI scans and 65 Equivalent tests were rendered using modeling tools for the tumor image [47]. (BRATS) Specific algorithms performed best for various sub-regions (performance attainment equivalent to human inter-rater variability), however, for all sub-regions at the same time no single algorithm ranked in the top. In whole tumor sensitivity and specificity were better in Zahao method than other methods, but performance of Bauer was better than others. This test can not Answer to question of which algorithm is "most beautifull" for segmentation of the glioma overall. They have figured out that for all three subtasks no single algorithm among the ones tested ranked in the top five. At the problems of "off-site" and "on-site," Hamamci and Geremia performed comparatively, Whereas the other functions performed in the "off-site" study were substantially better than in the prior "on-site" assessment. No single best known approach for all tumor areas recognized

2.3 Comparisons and Conclusions

Accuracy and sensitivity to the noise needs to go side to side in corner detection methods. The result of detection accuracy level is too much, then the noise in image may give false edge results. This cause the incorrect outlines. Also, if the noise degree is too extreme[24], parts of image outlines may become undetectable making objects unrecognizable.Therefore, edge detection methods should be used for images that are plain and that doesn't have noise.The reason is that, edge detection methods might detect wrong edges or extra unnecessary edges [30][11].

Segmentation methods are more simple and more sensitive to noise compared to the edge detection methods [24]. Perimeter-based approaches division the algorithm based on instantaneous shifts in strength close to edges while optimization methods such as area growth, watersheds and snakes divide the image into regions identical to a predetermined criteria set [11].

-Region growing methods are not usable for for validating segmentation; the problem with these methods are the tumor boundary segmentation consistency of the results. Such approaches are perfect for tumor homogeneity but not for tumor heterogeneity [31].

-Classification-based segmentation algorithms (K-NN, SVM, Bayers) accurately tumor segments and yield in large data sets it gives good results, but unwanted behaviors can occur where a class is not trained enough in the training stage[31].

-Clustering algorithms (K-means, Fuzzy, Hierarchical clustering) work very quick, fast and in images with no noise it gives good results but for images with noise it results in critical segmentation inaccuracies[31]. This can be achieved by using precise pre-processing algorithms (filtering, masking of the skull) before segmentation.

-Neural networks perform no better in the field of noise and no need to presume any fundamental allocation of data, but the greatest drawback is the learning process [31]. -BPN (Back Propagation Neural Network) node-increasing classifier provides quick and precise identification which can be used efficiently to segment high-precision MRI brain images, and the log sigmoid function is the best activation mechanism for medical image recognition applications[32].

-SVM can rank the brain image more accurately as normal or abnormal than the neural network SOM (Self-organizing maps)[32].
-Hybrid approaches are concluded to be better than using only one method. For Hybrid Segmentation BPN Neural network, SVM, Fuzzy, K-means can be used

3.Summary

3.1 Summary of Conceptual Solution

In this project, we want to implement a system that detects the tumor region and tumor type using image processing algorithms. There is a lot of experiments that does the same thing but we want to combine few of the past experiments and develop them considering their accuracy and computation time. We decided to use k-means algorithm because they are easy to implement. K-means algorithm is a segmentation algorithm which means they only detect the area/region that wants to be detected but they don’t necessarily classify the environment. Meaning segmentation algorithms only finds the different regions but does not define the region. We need to implement a classification algorithm that classy the area as tumor area and not tumor area. Mostly used and most successful classification methods are Neural Networks and SVMs. We decided to use SVM method to classify the region. We choose few articles as reference for implementation. We will firstly use preprocessing algorithms to normalize and reduce noise on image. Then we will implement k-means algorithm to detect tumor region and lastly we will use classification methods to classify the region. This methods and experiments are planned to develop to help doctors with their diagnoses. We also want to help doctors with early detection of tumors so that patients can start early detection.

These systems will be second hand judgement for their diagnoses.

3.2 Technology Used

We will use Python programming language and Opencv for the coding. Python works best with image processing since it also has special libraries we can use for image processing algorithms.

We will use Visual Studio Code and Jupyter editors.

We will use MySQL server for storing data because it works better with python programming language than other database servers.

There is no other technical tool needed for this project.

4. Software Requirements Specification

4.1 Introduction

4.1.1 Purpose

In past few decades, brain tumor segmentation in Magnetic Resonance Imaging (MRI) has become an emergent research area in the field of medical imaging systems. Accurate detection of brain tumor plays a vital role in the diagnosis of brain cancer. Our project aims to develop a program which analyses the Magnetic Resonance Images of the patient under consideration and recognizes the tumor by using image processing algorithms and detects the location of tumor. By using this method, patients can get information about their results .

Purpose of this document is to describe and give details about the project that planned to develop a method that detects brain tumor from MR Images. Research students, professors , teachers and all kind of people that wants to do exploration about ımage processing segmentation,artificial intelligence neural networks can benefit from this document. This document will also be helpful, for researchers who will do further projects about this area in the future and wants to improve the methods done so far.

4.1.2 Scope

This system involves determining the location and type of tumor in MRI images. The algorithms used today are not sufficient enough. We intend to use K-Means algorithm as segmentation method. If the algorithm shows the tumor region properly, we plan to use another method to detect tumor type. We want to show whether the tumor is cancerous or not. This way, the doctors will have second hand verification system for brain tumors. Algorithms will detect and highlight the tumor region area. If detection of tumor area is successful as planned, the method will pass on to second part of the project and will detect tumor type as cancerous or not. So, the second part of the project is optional. This proposed method/system will help the doctors in diagnosing brain tumors, and help the patients to start early medication/treatment.

4.1.3 Glossary

Medical Imaging: Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues

Image processing: Methods to perform to images to get images as output.

Image Segmentation: Image processing methods to divide image to segments using mathematical functions.

Neural Networks: Methods that designed to recognize patterns,modeled after human brain.

Hybrid Method: Using more than one image processing algorithms on an image.

4.3.4 Overview

In the first part of this document we gave a brief description about the project and explained why this document is necessary. In the second part, we gave much more detail about the project, so it can be more understandable for readers and users of this method. Third part is much more detailed; this part gives software details of the project. Third part is mostly for users, researchers who wants to use this method on their own. Third part explains and gives helpful guide for them to use this method own their own. We gave detail about functional and non-functional features of this developed method.

4.2 Overall Description

4.2.1 Product Perspective:

The proposed method planned to work on both Linux and Windows operating systems. The method will work on brain MR Images only. The method will use the developed image processing algorithms already exists. We plan to use more than one image processing algorithms for detecting brain tumor area. We will use preprocessing for reducing noise on MR Images and use segmentation algorithms for detecting tumor area. The users of this method should have knowledge about medical structure of brain tumor or image segmentation or both. The code will be sharable and portable to the public after the methodist published. The user can download the code on his/her computer and execute it for brain MR Images using any python editor.

4.2.2 User Characteristics:

It is proposed that doctors in this area to use the method or research student and teachers who are interested in image processing and want to develop a project about this can do experiments with this method. It is recommended that doctors who use this method for their patients to not give the result of the method anyone other than their patients to prevent patient privacy. Users of this system should only use MR brain Images. User should know python programming language if they want to improve or test the proposed method.

4.2.3 Constraints:

The system only accepts Brain MR Images, if another image is given the system gives an error and does not work correctly. The system only works with specific operating systems. The system is coded in Python language. This code might not work with future version of python programming.

4.2.4 Risks

We did a research about medical image processing before starting this project. We plan to choose some of these articles we read and implement them, and combine the ones that gives the best result. The risk is that the implementation of these methods can not be fast enough so we may not start the second part of this project. Also, we will use different datasets from the experiments done in articles so the algorithms and efficiency might give different results depending on datasets. Finding enough amount of datasets and finding clean, accurate datasets are the biggest obstacles for this project.

4.2.5 Assumptions and Dependencies

Future versions of Python may not work with current version of Python. The code of the method might have to change accordingly in the future.

4.3 Requirements

4.3.1 Specific Requirements

4.3.1.1 User Interfaces

There is no user interfaces for this project since there is no website or application will be developed. The project will be open source, and anyone who wants to access the code will be able to get Python codes. However, if the algorithm accuracy will be as expected and there will be enough time, a desktop application can be developed.

4.3.1.2 Hardware Interfaces

There is no hardware interface needed.

4.3.1.3 Software Interfaces

The codes will be available for both Linux and Windows. If the codes required for Linux and Windows work differently, two different codes for both Linux and Windows will be shared.

4.3.1.4 Communications Interfaces

No communication interface needed

4.3.2 Functional Requirements

User should be able to see the original MR Image,ımage after segmentation that shows the tumor region and image after detection tumor type.

The tumor region and type should be understandable for the user. The code should be easily readable and understandable for the user.

The system should detect tumor region accurately and should detect less than a minute.

The system should implement the algorithm properly and accordingly as; first preprocessing algorithms then segmentation algorithms.

4.3.2 Software system attributes

I. Reliability

The method will be published after, we make sure the algorithms are developed properly and the method is detection the area %70 accurately and fast enough as planned.

II.Availability

The method will be available for Linux and Windows operating systems. The code will be published and accessible for everyone to download.

III.Performance

The system should detect tumor region accurately and should detect less than a minute.

The system should implement the algorithm properly and accordingly as; first preprocessing algorithms then segmentation algorithms.

IV.Portability

The system will be useful in both Linux and Windows.

4.4. UML ANALYSIS MODEL

4.4.1 Use Case:

The following figure shows the use case diagram for the end users using the developed method. The most common users going to use the method will be doctors but any end-user can use our project. You can find our code on our github website. [64]. This use case diagram shows the steps all end-users can perform.

Detecting Cancerous Tumor Region Use Case Diagram:

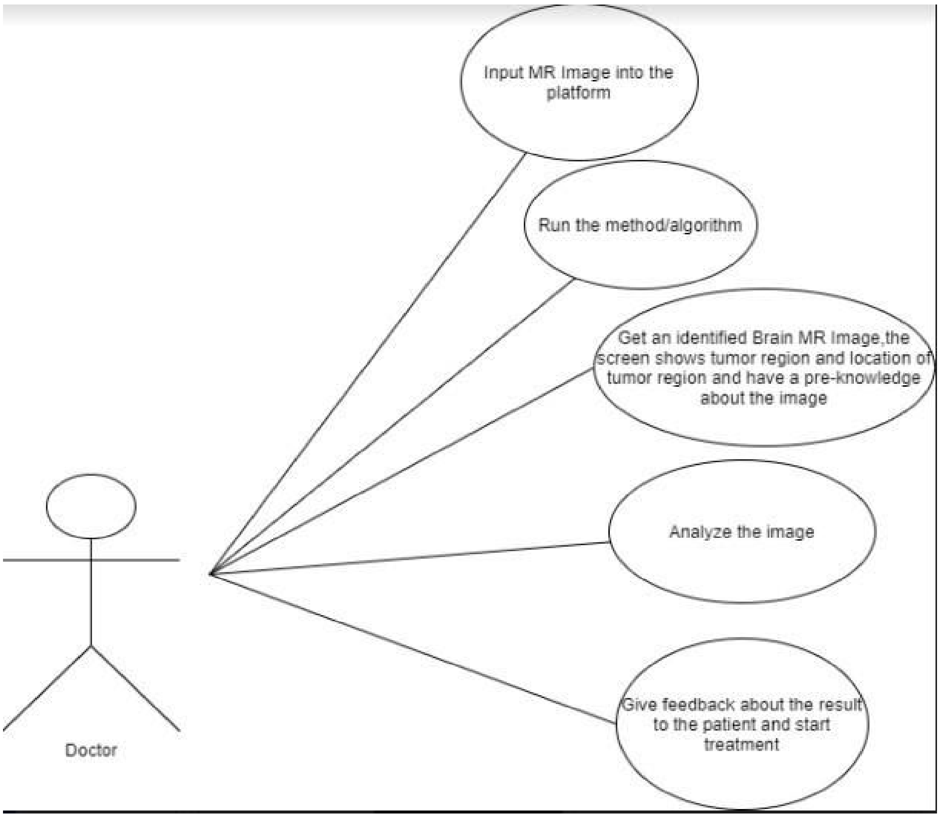

Figure 2: *Use Case Diagram 1 of proposed system*

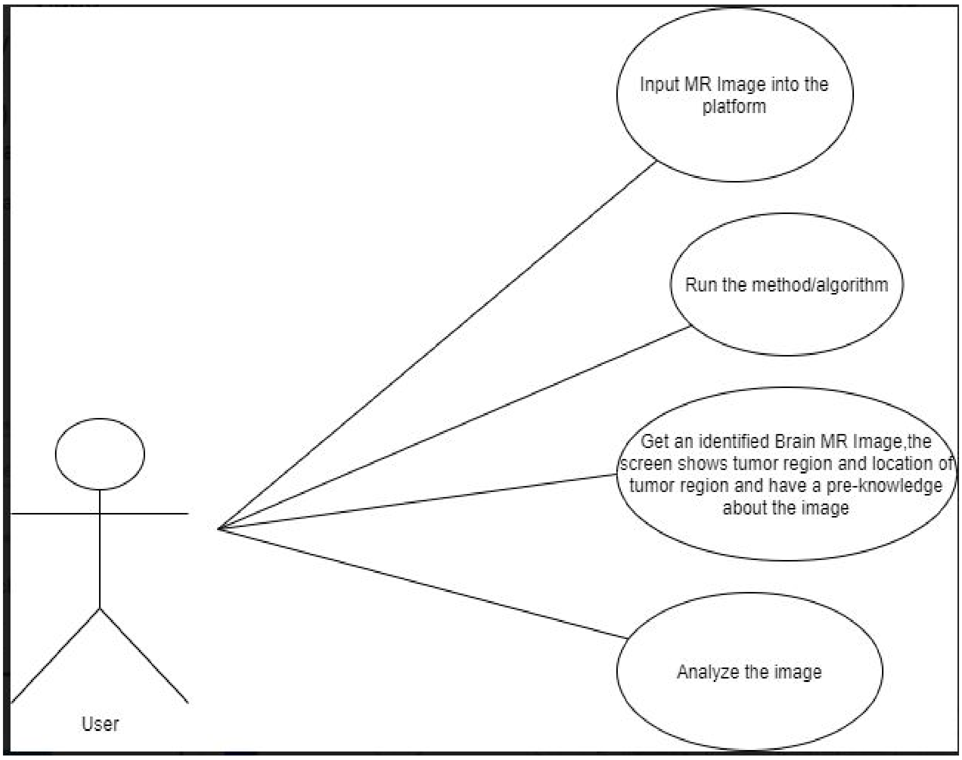

Figure 3: *Use Case Diagram 2 of proposed system*

4.4.1.1 Brief Description Of Use Case Diagram:

If a patient is suspected with a brain tumor and a brain MR Image of the patient is taken. Doctors can input the achieved MR Image into the system and run the algorithm developed from this project. After some time, the result of the algorithms and segmentation will be showed in the screen, this way doctor can have a pre-knowledge before doing diagnoses to the image, the developed methods can help and reduce time for doctor's diagnoses. If there is no tumor found from the algorithms the screen won't show any tumor area. If there is a tumor found from the algorithm; the screen will show the region of the tumor .The doctor will give the patient his/her result of the MR Image.

Overview	
Title	Detecting Cancerous Brain Tumor Area ın MR Image
Description	The doctor puts the MR Image of the patient to the system,systems shows that the MR Image has tumor area. The doctor analyze the image to be sure and diagnose the area as cancerous.The doctors talks to the patient and starts treatment immediately.
Actors and Interfaces	Doctor,
Initial States and Preconditions	The inputted image should be only brain image and should be an MR Image.
Basic Flow	
.1.Doctor ınputs ımage to the system 2.Run the Algorithm 3.After some little time,the screen shows an tumor area on the MR Image 4.Doctor analyzes the image 5.Doctor detects tumor area as cancerous 6.Doctor talks to the patient and starts the treatment	
Post Conditions	
None	
Alternative Flow(s)	
Detect Non-Cancerous Tumor Area: 1.Doctor ınputs the MR Image into the system 2.Doctor runs the algorithm 3.The system detects a tumor area 4.Doctor analyzes the image 5.Doctor detects tumor area as non-cancerous 6.Doctor talks to the patient and starts the treatment **Detect that there is no tumor:** 1.Doctors inputs the MR Image to the system 2.Doctors runs the algorithm 3.the system shows that there is no tumor 4.Doctor analyzes the Image	

Figure 4: *Detect Brain Tumor Doctor Use Case Description*

Overview	
Title	Detecting Cancerous Brain Tumor Area ın MR Image
Description	The user puts the MR Image of the patient to the system,systems shows that the MR Image has tumor area. The user analyze the system results.
Actors and Interfaces	User
Initial States and Preconditions	The inputted image should be only brain image and should be an MR Image.
Basic Flow	
1.User ınputs ımage to the system 2.Run the Algorithm 3.After some little time,the screen shows an tumor area on the MR Image 4.User analyzes the image 5.User takes necessary actions	
Post Conditions	
None	
Alternative Flow(s)	
Detect that there is no tumor: 1.User inputs the MR Image to the system 2.User runs the algorithm 3.the system shows that there is no tumor 4.User analyzes the results	

Figure 5: *Detect Brain Tumor End-User Use Case Description*

4.4.2 State Chart Diagram

State Chart of developed method:

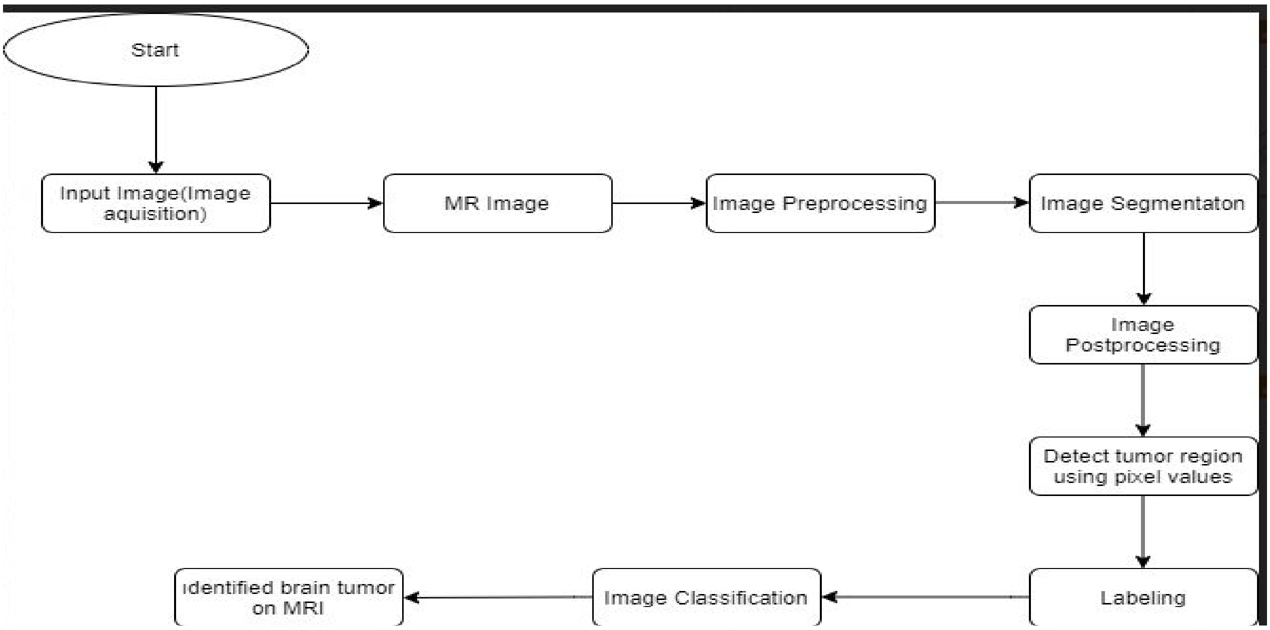

Figure 6: *State Chart of proposed system*

5. Software Design Description

5.1 Introduction

5.1.1 Purpose

Accurate detection of brain tumor plays a vital role in the diagnosis of brain cancer. In past decades many experiments has been done in medical ımage processing and artificial intelligence areas for detecting brain tumor, but no one algorithm or method outperformed other methods. There is no algorithm found that works and detects the location of brain tumor %100 accurately and has a fast computation time. So, there is no one superior method. Accuracy and computation time are 2 important components for medical algorithms. Since, brain tumor is critical and sensitive topic, detection of tumor needs to be done properly and correctly. Over these experiments with algorithms, it is found that neural network methods and hybrid methods gives the best results considering computation and accuracy.

Researchers , students ,professors and doctors who are interested in this field, wants to have more knowledge , wants to use or tests the method proposed can get benefits from this document.

5.1.2 Scope

For this project we want to work with medical images (MR Images) using image processing algorithms. We plan to use segmentation algorithm to detect whether brain tumor exists or not on brain MR Images. We plan to implement some of the methods used in past experiments in this field. We choose couple of experiments that gives the best results. We plan to implement these methods used on these experiments and compare their results depending on accuracy and computation time. We plan to combine these methods. From our research we chose few methods, we will combine these methods from the experiments. The combined methods are K-means and SVM. Combining them will have better results than using just one method. This is the first phase of the project. After first phase we will pass on to second phase and try to detect the type of the tumor as benign or malignant. We will also use preprocessing algorithms. MR Images carries a lot of noise which prevents the segmentation methods to work properly. These segmentation methods are used for detecting brain tumor region and noise decrease the efficiency of these methods working properly. So using preprocessing methods are necessary for this project. By using this method, doctors can get of their diagnosis verification and pre knowledge about the images. This project will also give contribution to medical image processing and artificial intelligence areas.

5.1.3 Glossary

Magnetic Resonance Imaging:

Magnetic resonance imaging is a method of medical imaging used during radiology to form representations of the anatomy and physiological processes of the body.

Brain Tumor Segmentation:

Segmentation of the brain tumor involves of distinguishing the various tumor tissues (active tumor, edema and necrosis) from normal brain tissues.

Image processing:

Image processing is a method of performing certain photo procedures to obtain an improved image or extract the useful information from it. It is a method of signal processing where input is an image and the output can be an image or features related to that image.

Image Segmentation:

Image segmentation is the procedure where an image is divided into multiple episodes. The segmentation of images is generally used for identifying points and boundaries in photos.

Neural Networks:

Neural networks are a group of algorithms which are broadly based on the human brain, developed to identify patterns.

Hybrid Method: Using more than one image processing algorithms on an image. Artificial Intelligence:

Intelligence shown by robots, in comparison to human natural intelligence.

5.1.4 Overview of document

This document gives technical details of the project planned to detect tumor region on MR Images. The first part of this document gives an overall description about the project. In the second part we gave detailed explanation of development of the project, design of the Project and planned Schedule for this project. In third part we gave technical details of the method we will develop.

5.1.5 Motivation

We are bunch of senior students in computer engineer department who are interested in image processing. Image processing has many fields. We want to improve ourselves on medical image processing. There is no, one algorithm that outperforms other algorithms on segmenting the medical images. Medical Image processing is a challenging and educational area, so we want to develop ourselves in this field. We also want to give good contribution to the world. Doctors can use this algorithms to be helpful for them on detecting brain tumors and start early treatment. This algorithms will be helpful to doctors. We did very large research in this field before starting the Project, we also studied python programming language. Python is the most suitable language for image processing since it has special libraries for image processing. Therefore, we will use python programming for this project.

5.2. DESIGN OVERVIEW

5.2.1 Description of Problem

There has been many image processing algorithms developed and many experiments done with these algorithms. Efficiency and accuracy of these algorithms are improved with these algorithms over the years. But, we still have long way to go to find the most successful algorithms. Detection of required area is sensitive and critical subject in segmenting medical images. Accuracy and fast computation time is two important scales for these segmentation algorithms. Hybrid methods and neural network methods gave the best results during these experiments but still there is no algorithm or method that leaves out other methods. We plan to implement some of these algorithms and combine them, if it's possible we also want to improve them. These methods will also help doctors in diagnosing brain tumors. Doctors will use these methods before diagnosing brain MR Images and will have pre knowledge about the image and also they will get a second judgment about the image. This way patients' diagnosis will be verified by a second person and the patients will start early treatment.

5.2.2 Technologies Used

We will use python programming language, we will use Opencv libraries. We plan to use Visual Studio Code and Jupyter editors. We will use MySQL Server for storing database. There is no other technical tool needed for this project.

5.2.3 Architecture Design

5.2.3.1 Sequence Diagram

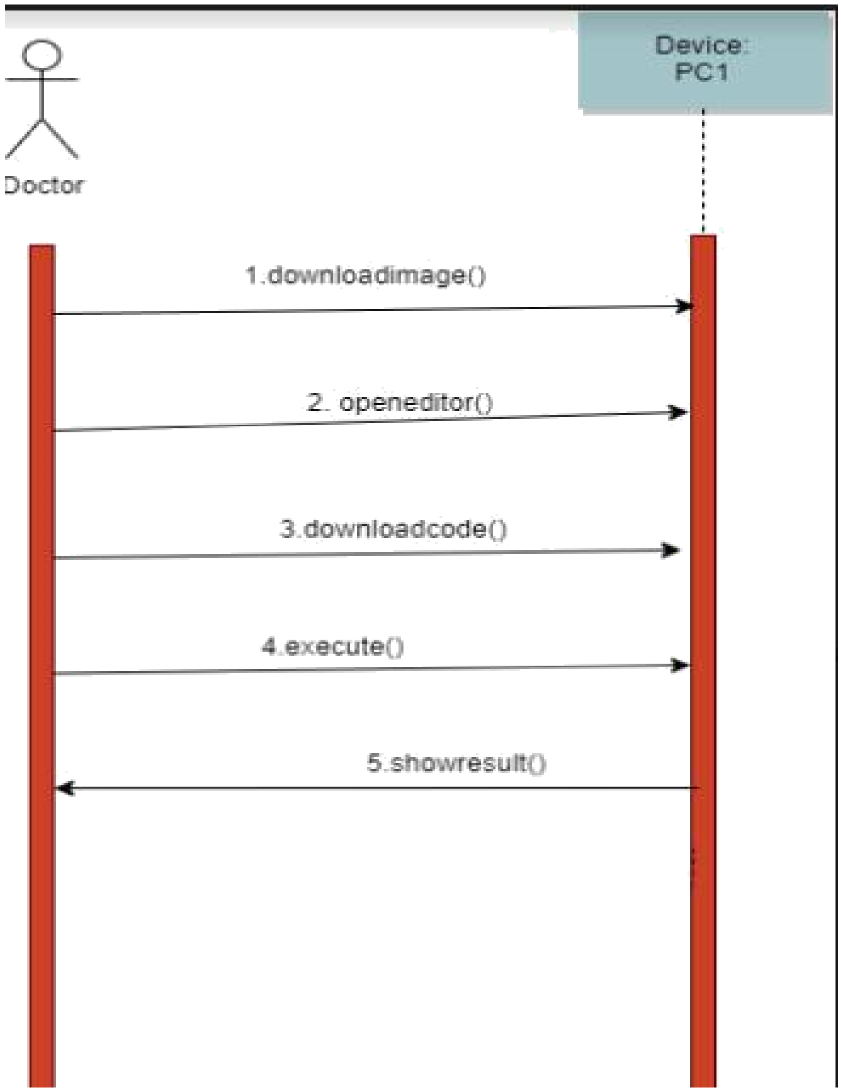

Figure 7: *Sequence Diagram of proposed system*

The Picture above shows the sequence diagram of the proposed method. The user needs to have an already existing brain MR image in their computer. The user then needs to download the code of our code and needs to open it in an editor that works with python programming.

The person then will execute the code and the system will show the results of the image. The system will show the original image and an image that is segmented (that shows the area of the tumor).

5.2.3.2 Class Diagram

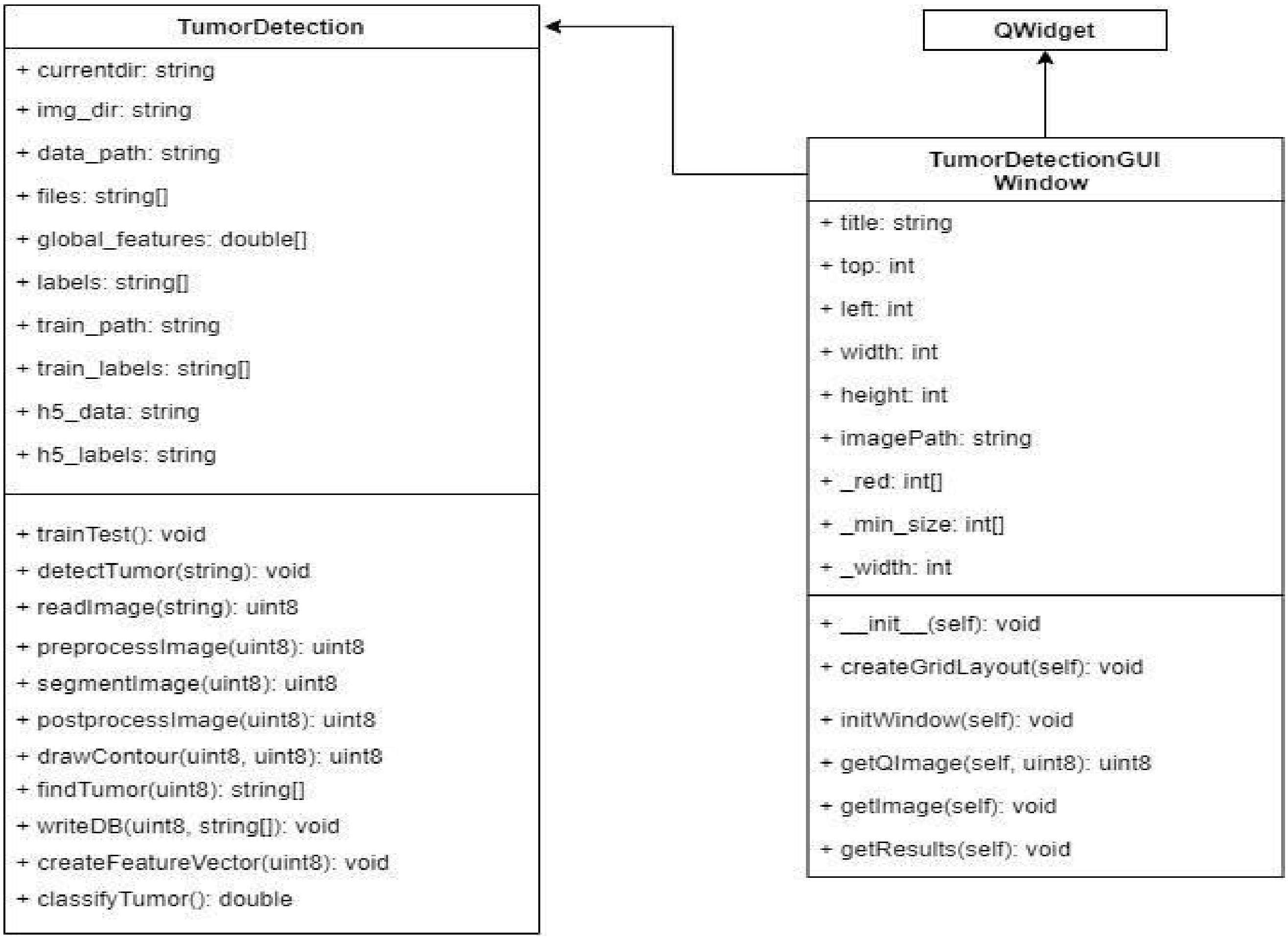

Figure 8: *Class Diagram of proposed system*

Class diagrams are used to show the classes, properties and relationships between each other in our project. There are 2 classes in our project. TumorDetection class is built on the image processing, segmentation and classification algorithms. TumorDetectionGUI class includes TumorDetection class and uses its functions to display the output of the algorithms to the end-user. We use a simple graphical user interface application on TumorDetectionGUI. In TumorDetection class: “currentdir” is a string that stores the user’s current directory that the user is executing the code on. “ımg_dir” and “data_path” are the directory of the whole dataset. “files” attribute is a string list that stores all datas in the dataset directory as a string. “global_features” attribute is a double list of features we are gonna use for training SVM. “labels” is also a list of strings. “labels” is for labels of images like if there is a tumor on the image we manually label the image as “yes” , if there is no tumor on the image we manually label it as “no”. “train_path” is a string attribute that we store in the dataset directory. “Train_labels” is a string list that we store labels of images that we use for training SVM. h5_data and h5_labels are string attributes that store the data result and result labels of SVM.

We firstly use the trainTest function. In trainTest we actually use detectTumor function. trainTest function reads the dataset file path from “file” string attribute and sends it to detectTumor function. detectTumor function first uses readImage function and sends the path to readImage function. readImage functions reads the image from path using opencv ready functions and returns the image. Then detectTumor function executes preprocessImage method. preprocessImage method takes the image as parameter. preprocessImage function turns the image to grayscale, does skull masking and then does median filter and returns the result image to detectTumor function. Then detectTumor function uses segmentImage function which takes preprocessed image as parameter. segmentImage function applies K-means clustering with k=3 returns resulted segmented image. Then detectTumor function uses postprocessImage function that takes segmented image as parameter. postprocessImage function calculates average pixel value of segmented image. If the image’s average pixel value is more than or equal to 120, we apply binary threshold with threshold value equal to average pixel value. If average pixel value is less than 120 apply binary threshold with threshold value 120. Then we use erode filter on segmented image. postprocessImage return the post processed image to detectTumor function and and detectTumor function executes drawContour function that takes both segmented image and post processed image. We find the contours of post processed image and draw the contours on segmented image. Contours are an outline representing or bounding the shape or form of something. We return segmented image with the contours on it to detectTumor function. Then detectTumor function uses findTumor function that takes post processed image as a parameter. findTumor function calculates total contour area of post processed image. If area is less than 0 we label it as “no”, else we divide pixel values from the middle of the image and calculate total left and right pixel values. We label image as yes. We have second string list attribute we use for labeling images that has tumor on it called status. If append “left” or “right” based on the left and right pixel values. findTumor functions return status label list to detectTumor function. Then detectTumor function executes writeDB function. writeDB function stores dataset images and labels on MySQL database. Then detectTumor function executes createFutureVector function that takes original image as parameter. On createFeatureVector we calculate HU Moments of image and append it on global_features list. HU Moments are used for to detect attributes of images and to train SVM with this attributes. Lastly we execute classifyTumor function which splits dataset as train and test. We apply SVM on a dataset using labels and feature vector (global vector) that stores attributes of images . Then we apply K-fold cross validation. K-fold cross validation is used for evaluating SVM accuracy. Then with K-fold cross validation we calculate prediction accuracy of SVM.

The purpose of TumorDetectionGUI code is that an end-user can load a brain MR image and get the results of segmentation algorithms in TumorDetection class visually. There is a Window class that uses Qt framework for Python applications. There are functionalities that Qt provides such as signals & slots mechanism. Those functionalities help to catch button events and process data according to that. This class is inherited from QObject and built on QApplication.

For creating a desktop application, we need to define main window properties, and those properties in this class are title, top, left, width, height, _red, _width and _min_size. There is an imagePath variable to take a path from user when he/she loads an image and pass it to the TumorDetection class for processing an image.

First of all, system starts creating a Window object and its constructor is __init__ method. In this method, window properties is set and initWindow method is called. The main job initWindow is to initialize buttons, labels, layout and set those things to the application window. While initializing a layout, createGridLayout function is called to set all buttons and labels in a grid layout. After that, in order to catch an action event such as clicking a button, the signals (button clicked) are connected to the related slots. There are two slots to handle button clicking operation. The first one is getImage slot and it is responsible from creating a file dialog to take image path, then send it to readImage method in TumorDetection class, and display read image on window. However, since TumorDetection reads an image in OpenCV format, showing it on the window requires converting this image to QImage format, and getQImage method does this operation. The second slot handles results of the original image. If the user pushes "Results" button, its signal is triggered and getResults slot will be operated. This slot is responsible from calling each related function in TumorDetection class and set the results of those functions to the application window. Finally, the application window is shown to the user.

In the next chapter we show flowchart of the system. You can check the flow-chart diagram for better understanding.

5.2.3.3 Flow-chart Diagram

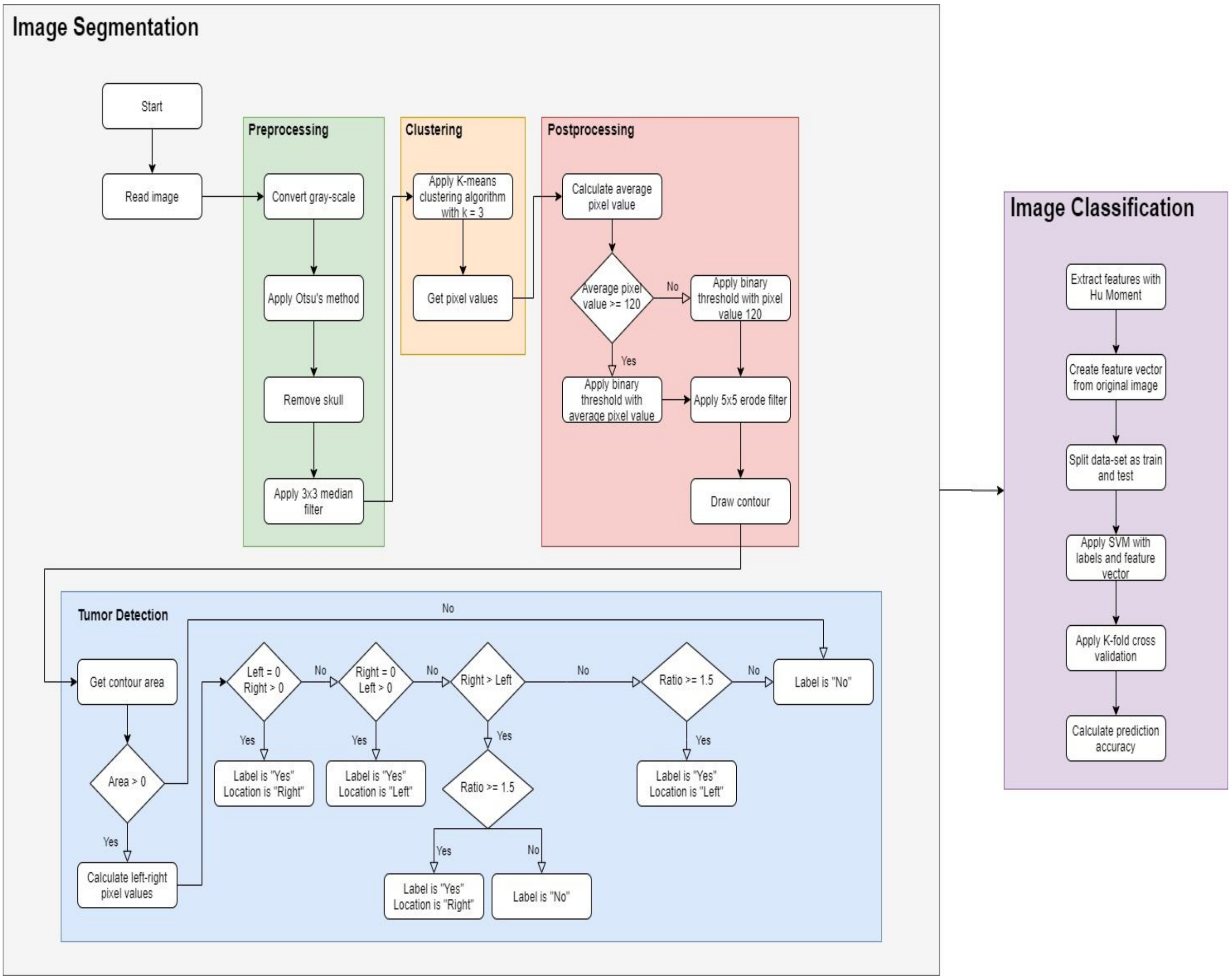

Figure 9: *Flow-Chart Diagram of proposed system*

We first used preprocessing methods on images and those are: converting to grayscale, skull masking, and median filtering respectively. After that, the segmentation method that is K-means clustering with k=3 has been applied to images. Then we applied post processing methods. We calculated average pixel values of the image. If the image's average pixel value is more than or equal to 120, we apply a binary threshold with threshold value equal to average pixel value. If the average pixel value is less than 120 apply binary threshold with threshold value 120. Then we apply an erode filter to images.After that we draw contours of image.Then we calculate contour area of images if area is less than 0 we label it as “No”. Else we calculate the total left and right pixel value of images.If left pixel values are equal to 0 and right pixel values are more than 0, we label image as “yes” and label status as “right”. If right pixel values are equal to 0 and left pixel values are more than 0 we label it as “yes” and we label status as “left”. If any of them is not 0, we check if right pixel values are bigger than left pixel values with ratio 1.5. If ratio is equal or bigger than 1.5 we label image as “yes” and label status as “right”. If ratio is less than 1.5 we label the image as “no”. If left values are bigger than right values and ant of them is not 0 we check the 1.5 ratio again. If ratio is bigger or equal to 1.5 we label it as “yes” and label status as “left” else we label it as “no”. Then we go to Tumor Classification stage where we firstly extract image features with calculating HU Moments of images. Then we create feature vector from original images. Then we split our dataset as train and test dataset. We apply SVM with labels and feature vector. Then we apply K-fold cross validation to calculate accuracy of SVM. Results have been stored in the database by using the MySQL tool.

5.2.3.4 Project Plan

DATE	21.10.2019 - 01.11.2019	02.11.2019 - 17.11.2019	17.11.2019 - 22.11.2019	23.11.2019 - 13.12.2019	13.12.2019 - 20.12.2019	15.01.2020 - 31.02.2020	01.03.2020 - 13.03.2020	14.03.2020 - 20.03.2020	21.03.2020 - 20.05.2020	25.04.2020 - 30.04.2020	01.05.2020 - 08.05.2020	09.05.2020 - 15.05.2020	16.05.2020 - 29.05.2020
WORK													
Researching related work Writing literature review document	TEAM												
Designing web page for project Writing SRS document		TEAM											
Finishing last changes of SRS document Starting research on SDD document			TEAM										
Writing SDD document				TEAM									
Project report Project presentation					TEAM								
Coding						Nisan Çandır Ece Yılmaz							
Preparing project work plan							TEAM						
Test plan document								TEAM					
User manual										Deniz Yıldız Kevser Sürgüveç			
Project report											TEAM		
Project poster												TEAM	
Last improvements													TEAM

Figure 10:*Table of Work Plan*

Table 1: *Detailed information of work plan*

Work	Description	Assignee(s)
Writing literature review document	Researching related literature, reading and documenting articles about the algorithms that will be used in the project	Team
Designing web page for project	Designing web page for publishing documents about project and team information	Team
Writing SRS document	Determining and documenting software requirements for project	Team

Work	Description	Assignee(s)
Writing SDD document	Identifying how data design and architecture design of software will be, documenting design description	Team
Project report	Collecting all documents about project in a single document	Team
Project presentation	Preparing presentation slide and making a demo presentation about what has been done so far	Team
Coding	Developing/ coding the system on development environment	Nisan Çandır Ece Yılmaz
Preparing project work plan	Deciding on a work plan for the rest of the semester	Team
Test plan document	Prepare test plan document considering important test cases and test criterias.	Team
User Manual	Write user manual document explaining end-users how to set up and use the system	Kevser Sürgüveç Deniz Yıldız
Project Report	Go over project report and add missing parts	Team
Project Poster	Prepare the project poster that represents the project	Team
Last Improvements	Make the last improvements and changes on the project.	Team

5.3. DATA DESIGN

5.3.1 Data Description

5.3.1.1 Entity Relationship (ER) Model

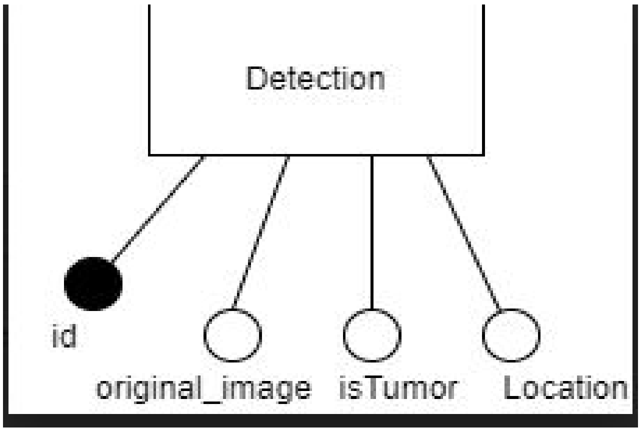

***Figure 11**: ER Model of proposed system*

In the Figure, we showed the Image Analysis for the Classification of Brain Tumor Location on MR Images Entity Relationship Diagram.We use only one table. Id is the primary key of the table. We store original image as binary on the database. isTumor attribute is a bit that shows if there is a tumor on the image or not. If there is a tumor it is 1 if there is not it is 0. Location is a String and it shows the location of tumor on the image as left or right. If there is no tumor (İsTumor=0) Location is empty (Location=" ")

We will use MySQL database server for storing database, since it is easy to use with python programming.

5.3.2 Data

The data to be used on the system should be brain MR images. MR image types are usually dicom and we expect to use dicom MR Images. Dıcom images usually store the tissue of the image, type of the image and information about the patient in its format.

DICOM (Digital Imaging and Communications in Medicine) is for handling, processing, printing and conveying medical imaging information. This includes a definition of the file format and a network communication protocol [1]. We plan to get the MR images to use on the system and to implement algorithms from TCIA's [2], BRATS's [3] and Kaggle's [4] websites.

5.4. USE CASE REALIZATION

5.4.1 Project Components

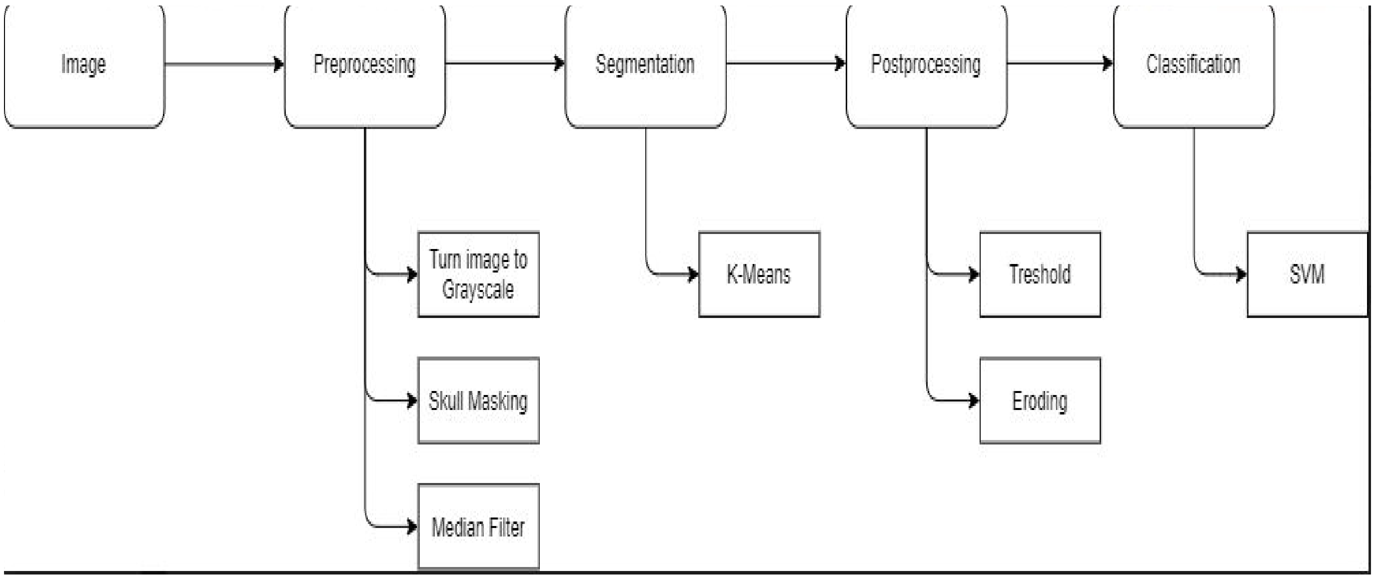

Figure 12: *Block Diagram of Project Components*

5.4.1.1 Brief Description of Project Components

From the research we did before starting this project, we picked 3 articles as references to implement for our project. These articles explain the details very well and very understandable and explains easy to implement algorithms.. Therefore, we decided to use K-means and SVM. We practiced Fuzzy nd K-means method, and we found out that Fuzzy doesn't give good results. As a result, we used only K-means as a segmentation method and SVM afterwards. We also use contours methods to find tumor area and use labeling, calculate HU Moments to find features of images before SVM Method. These calculations are necessary for the SVM method. We don't mention them here because they are the details of project components. We only want to give general information about components.

The articles we choose as references are: [9] [10] [11].

5.4.1.1.1 Ready Packages

Ready packages in Python which will be used for machine learning algorithms (clustering, classification) and for image processing are listed as below:

- Numpy: for multi-dimensional array and matrix processing (image).

- Scipy: for optimization, linear algebra, integration and statistics.
- Scikit-learn: for machine learning algorithms, supports most of the supervised and unsupervised learning algorithms (clustering and classification).
- Matplotlib: for data visualization.
- Pydicom: for dealing with MRI images.

5.4.1.1.2 Image Preprocessing

Preprocessing is a common name for image operations at the lowest abstraction level; all input and output are images of the strength.Preprocessing aims to improve image data which suppresses unwanted distortions or enhances certain image features which are necessary for further processing (segmentation).[5] In this project, preprocessing algorithms are used to reduce noise and distortions so that, segmentation algorithms can detect tumor region more accurately. We plan to use 3 preprocessing algorithms for this project; Median Filter, High Pass Filter, Skull Masking Median Filter. The Median Filter is a technique that is not linear and is often used to eliminate noise from an image or signal.

Such noise reduction is a standard preprocessing phase aimed at improving the later processing performance.It is particularly effective at removing noise of the form ' salt and pepper. ' The media filter works by moving pixel by pixel through the image, replacing each value with the neighboring pixel mean value. The neighbors pattern is called the "window" that slides, pixel by pixel over the whole image. The median is determined by first sorting all the pixel values in numerical order from the window, and then replacing the pixel considered with the middle (median) pixel. [6]

5.4.1.1.3 Skull Masking

MR brain imaging studies also involve a specific processing to distinguish the brain from extra-cranial or non-brain tissue through MRI head scans, commonly known to as skull stripping. because the brain images that have been pre-processed with automated skull stripping eventually lead to better segmentation of various regions of the brain leading in accurate diagnosis of different brain-related diseases. Skull stripping is a basic step designed to eliminate non-brain tissue from MR brain images for many clinical applications and examinations, its accuracy and speed are considered as critical factors in the segmentation and analysis of brain images.Efficient and automated procedures for skull stripping help increase the speed and accuracy of diagnostic procedures in medical applications. [8]

5.4.1.1.4 Image Segmentation

Image segmentation is known as an critical feature of digital image processing to divide the image into different sections and regions.The result of image segmentation is a collection of portions that contain the entire image or a series of outlines taken from the image overall. [12]

K-Means

K-means is the simplest unsupervised learning algorithm that can overcome segmentation challenges.The method followed is very simple for identifying a given set of data through a certain amount of clusters. Centers are defined in K-means ' K, ' per each cluster.

Such clusters must all be placed away from each other. The next step is getting a point that belongs to a given set of data and associate it with the closest center.

The first phase is completed when no stage is pending, and early grouping is done The second step is to recalculate the ' k ' new centroids resulted from the previous step as the barycenter o f the clusters. After having new centroids with ' K ' a new connection must be made between the same data set points and the nearest new center.

A loop was developed. As a result of this process, the centers of k change their location step b y step until the centers are no more moving. [10]

5.4.1.1.5 Post Processing

We used postprocessing methods after K-means to increase accuracy of SVM method. With these methods the area of tumor is more apparent.

Threshold

Threshold is a image processing method. Chosen a pixel value, we set pixel values to upper or lower bound comparing to the threshold value. In our project the upper and lower values are 0 and 1. Our threshold value is 120. If a pixel value is equal to 120 or higher than 120, the new value of that pixel is 1. If pixel value is lower than 120, new pixel value is 0.

Eroding

Erode is image processing method used to reduce area of shapes. We used this method to reduce white noises in the image which mistakenly seems like tumor area.

5.4.1.1.6 Image Classification

Image classification corresponds to a computer vision method that is able to classify an image according to its visual content. For instance, an image classification algorithm can be intended to tell whether or not an image includes a female figure.[13]

5.4.1.1.7 SVM

Rather than just reducing an objective feature centered on the training , the SVM tries to reduce the limitation on the generalization fault.

Therefor, when implemented to data outside the training set an SVM tends to perform better. SVM achieves that advantage by concentrating on its most challenging to identify training examples. These examples of "borderline" training are called support vectors.

The SVM classifier is centered on the hyperplane that maximizes the margin of separation between the two groups..

SVM considers a collection of input data and predicts which of two possible classes forms the output for each given input, allowing it to be a non-
probabilistic linear binary classifier. SVMs also can perform an efficient non- linear classification through kernel technique, mapping their inputs into high- dimensional feature spaces implicitly.
To optimize the margin among groups, the theoretical risk and classification is managed using its kernel.

SVM has two phases; training and testing. SVM trains itself through attributes that its learning method is given as an input. SVM selects the appropriate margins among two classes when training. Attributes are marked with a specific class according to the class associative.

Artificial neural network has a couple of problems with local minima and amount of selected neurons for each issue. To address this issue, SVM does not take local minima and the selection of neurons by introducing the idea of hyperplanes [10].

6. Test Cases:

6.1. Introduction

6.1.1 Version Control

Table 2: *Table of product versions and release dates*

Version No	Description of Changes	Date
1.0	First Version	April 24, 2020

6.1.2 Overview

In this project we are developing a system that detects tumor region from brain MR images. The functions/methods that are used in the system will be tested if they are working correctly and giving good results. SVM method will be tested if it is giving correct classification results.

6.1.3 Scope

This document encapsulates the test plan and test design specifications.

6.1.4 Terminology

Table 3: *Table of methods to be tested*

Acronym	Definition
Preprocessing Methods	Preprocessing methods are used to reduce noise and to improve image features so that higher methods can work more accurately.
Segmentation Methods	Segmentation methods are used to divide brain in regions so that we can separate a tumor from other regions.
Classification Methods	Classification methods are used to classify tumor status whether there is a tumor or not automatically according to features of given input image. Segmentation methods only separates tumor region from others but does not know what tumor is.

6.2 FEATURES TO BE TESTED

This section lists and gives a brief description of all the major features to be tested.

6.2.1 Preprocessing Methods

We use three major different methods for preprocessing those are thresholding, median filtering and skull stripping. This part includes testing the result of each methods if they will give good results.

6.2.2 Segmentation Methods

We use K-means algorithm for image segmentation. This part includes testing results of clustering method whether object labeling after the segmentation is successful or not.

6.2.3 Classification Method

We use SVM as a classification method. Classification methods need two different datasets one is for training the data, and the other one is for testing after training is done. Checking/testing if classification method gives accurate results because training is a sensitive, critical topic.

6.3 FEATURES NOT TO BE TESTED

We will not test the amount of time each function/method takes. Since in this project we do not plan to improve the methods we plan to develop a new system with already existing methods.

6.4. ITEM PASS / FAIL CRITERIA

6.4.1. Exit Criteria

- 100% of the test cases are executed.
- 95% of the test cases passed.
- Segmentation method gives at least 70% accuracy.
- Classification method gives at least 60% accuracy since it depends on the accuracy of the segmentation algorithm.

6.5 Test Cases

Here list all the related test cases for this feature

Table 4: *Table of test cases in detail*

TC ID	Requirements	Priority	Scenario Description
IAFTCOBTLOMI.T.01	3.1.1	H	User can upload images
IAFTCOBTLOMI.T.02	3.1.1	H	User can see outputs of images
IAFTCOBTLOMI.T.03	3.2.1	H	System gives output on preprocessing algorithms.
IAFTCOBTLOMI.T.04	3.2.1	H	System gives output on segmentation algorithm.
IAFTCOBTLOMI.T.05	3.2.1	H	System gives output on classification algorithm.

7.Test Results

7.1. Individual Test Results

Table 5:*Test Results Table*

TC ID	Priority	Date Run	Result	Explanation
AFTCOBTLOMI.T.01	H	25.04.2020	Pass	The user can upload a brain mr image that he/she wants to use
AFTCOBTLOMI.T.02	H	25.04.2020	Pass	The user can see the result ima and whether or not it has tumor.
AFTCOBTLOMI.T.03	H	25.04.2020	Pass	Preprocessing algorithms work accurate and gives outputs
AFTCOBTLOMI.T.04	H	25.04.2020	Pass	Segmentation algorithm work accurate and gives results
AFTCOBTLOMI.T.05	H	25.04.2020	Pass	Classification algorithm work accurate and gives results.

7.2 Summary of Test Results:

Table 6: *Summary of Test Results Table*

Priority	Number of TCs	Executed	Passed
H	5	20	20
M	0	0	0
L	0	0	0
Total	5	20	20

All high priority test cases passed our test cases.We have executes over 20 test cases and our test cases passes all of them.Exit criteria is met.

7.3. Exit Criteria

Table 7: *Table of criteria and whether they met on not*

Criteria	Met or Not
100% of the test cases are executed	M
95% of the test cases passed	M
Segmentation method gives at least 70% accuracy.	M
Classification method gives at least 60% accuracy since it depends on the accuracy of the segmentation algorithm.	M

We have executed all test cases and 100% of test cases are passed. Also, 100% of high priority test cases are passed Exit criteria is met.

8. Conclusions

We learned details of image processing algorithms that used on medicine that we didn’t know before. We learned mathematical details of these algorithms. We also had pre-knowledge of implementing these algorithms on python programming language. We gained more general knowledge about image processing field. We did a very broad research about the project. We read a lot of articles that detects brain tumor region on MRI Images using Image Processing methods. We planned the software requirements of the project and activity diagram of the project. We also planned the design of the project such as class diagram and database diagram. This way, we planned the project before starting the implementation to prevent any crisis and misleading. We plan to combine more than one experiment, this way the system will give better results than using just one method. We plan to develop the methods used in these experiments. The biggest risks of this project is finding accurate and enough dataset. Also we might not be able to develop and implement these methods properly. This is also one of the biggest risk of the project. This academic project will have contributions on image processing on medicine applications and future works. This project will also help doctors on detecting tumors on early stages.

9. References

[1] Kekre, H. B., Tanuja K. Sarode., Saylee M. Gharge (2009):"Detection and Demarcation of Tumor using Vector Quantization in MRI images", IJEST, Vol: 1, No: 2, pp: 59-66.

[2] MRI Basics- https://my-ms.org/mri_basics.htm

[3] T. Kalaiselvi, P. Nagaraja, “Brain Tumor Segmentation from MRI scans through Bit-plane Slicing and Clustering Techniques”, In the Proceedings of the 2015 National Conference on Recent Advances in Computer Science and Applications (NCRACSA2015), Gandhigram, India, pp.89-94, 2015.

[4] B. Menze, A. Jakab , S. Bauer, J.K. Cramer, K. Farahani K, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. IEEE Transactions on Medidal Imaging, Vol. 34, pp.1993-2024, 2015

[5] S.V. Aswathy, G. Glan Deva Dhas and S.S. Kumar, “A Survey on Detection of brain tumor from MRI brain images”, IEEE International Conference on control, Instrumentation, Communication and computational technologies (ICCICCT), 2014.

[6] Luxit Kapoor and Sanjeev Thakur, “A survey on brain tumor detection using image processing techniques”, IEEE international conference on cloud computing, data science,and engineering confluence, pp. 582-585, 2017.

[7] Rajesh C.Patel and A. S. Bhalchandra, “Brain tumor extraction from MR Images using Matlab”, International Journal of electronic communication and soft computing science and engineering, vol. 2 no. 1, 2012.

[8] J. Sleva Kumar, A. Lekshmi and T. Arivoli, ”Brain tumor segmentation and its area calculation in brain MR Images using K means clustering and fuzzy C means algorithm”, IEEE international conference on advances in engineering science and management (ICAESM), 2012.

[9] M. N. Wu, C.C Lin and C.C. Chang, “Brain tumor detection using color-based K means clustering segmentation”, Third International conference on intelligent information hiding and multimedia signal processing (IIH-MSP), 2007.

[10] N.N Gopal and M. Karnal, Diagnose brain tumor through MRI using image processingclustering algorithms such as fussy C means along with intelligent optimization techniques”, IEEE International conference on computational intelligence and computing research, 2010

[11] MRI BRAIN IMAGES SEGMENTATION TECHNIQUES – A REVİEW, D.SELVARAJ Department of ECE, Sathyabama University, Chennai, Tamilnadu, India R.DHANASEKARAN, Syed Ammal Engineering College, Ramanathapuram, Tamilnadu, India, D.Selvaraj et.al / Indian Journal of Computer Science and Engineering (IJCSE), Vol. 4 No.5 Oct-Nov 2013

[12] TCIA(The Cancer Imaging Archieve) website - https://www.cancerimagingarchive.net/

[13] https://www.cse.unr.edu/~bebis/CS791E/Notes/RegionGrowing.pdf

[14] Mancas, M., Gosselin, B., Macq, B(2005): “Segmentation using a Region Growing Thresholding,” Proceedings of the Electronic imaging Conference of the International society for optical imaging (SPIE/EI 2005), San Jose.

[15] Dzung L.Pharm, Chenyang Xu, Jerry L.Prince (1998): “A Survey of Current Methods in Medical Image Segmentation,” Technical Report JHU / ECE 99-01, Department of Electrical and Computer Engineering.

[16] Brinkmann, B.H., Manduca, A., Robb, R.A(1998): “Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction,” IEEE T. Med. Imag., 17:161–171.

[17] Warfield, S., Duda, (2001): “K-Nearest Neighbour Classification”.

[18] El-Sayed, A., El-Dahshan, Abdel-Badeeh, M. S. Tamer, H. Y (2010): “A Hybrid Technique for Automatic MRI Brain Images Classification”, digital Signal Processing, Vol. 20, Issue 2, March 2010, pp. 433-441.

[9] M. N. Wu, C.C Lin and C.C. Chang, “Brain tumor detection using color-based K means clustering segmentation”, Third International conference on intelligent information hiding and multimedia signal processing (IIH-MSP), 2007.

[10] N.N Gopal and M. Karnal, Diagnose brain tumor through MRI using image processing clustering algorithms such as fussy C means along with intelligent optimization techniques”, IEEE International conference on computational intelligence and computing research, 2010

[11] MRI BRAIN IMAGES SEGMENTATION TECHNIQUES – A REVİEW, D.SELVARAJ Department of ECE, Sathyabama University, Chennai, Tamilnadu, India R.DHANASEKARAN, Syed Ammal Engineering College, Ramanathapuram, Tamilnadu, India, D.Selvaraj et.al / Indian Journal of Computer Science and Engineering (IJCSE), Vol. 4 No.5 Oct-Nov 2013

[12] TCIA(The Cancer Imaging Archieve) website - https://www.cancerimagingarchive.net/

[13] https://www.cse.unr.edu/~bebis/CS791E/Notes/RegionGrowing.pdf

[14] Mancas, M., Gosselin, B., Macq, B(2005): “Segmentation using a Region Growing Thresholding,” Proceedings of the Electronic imaging Conference of the International society for optical imaging (SPIE/EI 2005), San Jose.

[15] Dzung L.Pharm, Chenyang Xu, Jerry L.Prince (1998): “A Survey of Current Methods in Medical Image Segmentation,” Technical Report JHU / ECE 99-01, Department of Electrical and Computer Engineering.

[16] Brinkmann, B.H., Manduca, A., Robb, R.A(1998): “Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction,” IEEE T. Med. Imag., 17:161–171.

[17] Warfield, S., Duda, (2001): “K-Nearest Neighbour Classification”.

[18] El-Sayed, A., El-Dahshan, Abdel-Badeeh, M. S. Tamer, H. Y (2010): “A Hybrid Technique for Automatic MRI Brain Images Classification”, digital Signal Processing, Vol. 20, Issue 2, March 2010, pp. 433-441.

[19] Clarke, L. P., Velthuizen, R. P., Phuphanich, S., Schellenberg, J. D., Arrington, J. A. Silbiger, M (1993): "MRI: Stability of Three Supervised Segmentation Techniques", Magnetic Resonance Imaging, 11: pp. 95-106, 1993.

[20] Zhang, K., CAO, H.X., Yan, H (2006): "Application of support vector machines on network abnormal intrusion detection", Application Research of Computers, Vol.5, pp.98-100.

[21] Pham, D., Xu, C. Prince, L. (2000):"Current methods in medical imagesegmentation", Annual review of Biomedical Engineering, PP. 315-337.

[22] Cox, E.,(2005): " Fuzzy Modeling and Genetic Algorithms for Data mining and Exploration,"
Elsevier, 2005

[23] Hall, L.O., Bensaid, A.M., Clarke, L.P., Velthuizen, R.P., Silbiger, M.S., Bezdek. J (1992): "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," IEEE Trans Neural Network, Vol.3, pp.672–682

[24] Kang, W.X., Yang, Q.Q., Liang, R.R (2009): "The Comparative Research on Image Segmentation Algorithms", IEEE Conference on ETCS, pp. 703-707.

[25] Zhu, C., Ni, C., Li, Y., Gu, G (2009): "General Tendencies in Segmentation of Medical Ultrasound Images", International Conference on ICICSE, pp. 113-117.

[26] Rumelhard, D., Mcclelland, J,(1986): "Parallel Distributed Processing", MIT Press, Cambridge, Mass.

[27] Jabbar, N., Mehrotra, M (2008): "Application of fuzzy neural network for image tumor description," Proc. of World Academy of Science, Engineering and Technology; 34:575-77.

[28] Chapter 1 Hybrid Segmentation Methods - https://academiccommons.columbia.edu/doi/10.7916/D8D226WB/download

[29] Gonzalez, Rafael C., Richard E.Woods (2007): "Digital Image Processing", Pearson Education.,
3rd Edition.

[30] Varshney, S. S., Rajpal, N., Purwar, R (2009): "Comparative Study of Image Segmentation Techniques and Object Matching using Segmentation", Proceeding of International Conference on Methods and Models in Computer Science, pp. 1-6.

[31] A Review on Automated Brain Tumor Detection and Segmentation from MRI of Brain, Sudipta Roy, Sanjay Nag, Indra Kanta Maitra, Samir Kumar Bandyopadhyay, 2013

[32] A REVIEW ON CURRENT MRI BRAIN TISSUE SEGMENTATION, FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES, D. SELVARAJ, R. DHANASEKARAN,2013

[33] Brain Tumor Segmentation from MRI Head Scans through GSO based FCM Clustering and Region Growing Technique, P.Nagaraja1, T.Kalaiselvi, P.Sriramakrishnan, JCSE,Vol-6, Special Issue-4, May 2018

[35] Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network,Cui S,Mao L,Jiang JLiu C, Xiong S,J Healthc Eng, 2018 Mar

[36] S. R. Telrandhe, A. Pimpalkar, A. Kendhe (2016). Detection of Brain Tumor from MRI images by using Segmentation & SVM.

[37] L. Zhao, K. Jia (2016). Multiscale CNNs for Brain Tumor Segmentation and Diagnosis.

[38] Z. Kap´as, L. Lefkovits, L. Szil´agyi (2016). Automatic Detection and Segmentation of Brain Tumor Using Random Forest Approach.

[39] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino (2016). A Hybrid Feature Selection With Ensemble Classification for Imbalanced Healthcare Data: A Case Study for Brain Tumor Diagnosis.

[40] T. Gupta, P. Manocha, T. Gandhi, R. Gupta, B. Panigrahi (2017). Tumor Classification and Segmentation of MR Brain Images

[41] S. Kumar, C. Dabas, S. Godara (2017). Classification of Brain MRI Tumor Images: A Hybrid Approach.

[42] C. Rao, P. Naganjaneyulu, K. Prasad (2017). Brain Tumor Detection and Segmentation Using Conditional Random Field

[43] N. Bahadure, A. Ray, H. Thethi (2017). Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM

[44] P.M. Shakeel, T.E. El. Tobely, G. Manogaran, H. Al-Feel (2019). Neural Network Based Brain TumoR Detection Using Wireless Infrared Imaging Sensor.

[45] M.S.Alam, Md Mahbubur Rahman, Mohammad Amazad Hossain, Md Khairul Islam, Kazi Mowdud Ahmed, Khandaker Takdir Ahmed, Bikash Chandra Singh, Md Sipon Miah (2019). Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm .

[46] T. Logeswari, M. Karnan (2019). An Improved Implementation of Brain Tumor Detection Using Segmentation Based on Soft Computing.

[47] Mukambika P. S., Uma Rani K. IRJET. Segmentation and Classification of MRI Brain Tumor

[48] Mamidi Srujana , Mudadla Chandini , Ijjirythu Lavanya , R. Divyakanthi. Tumor Identificationin Brain MRI Images Using Fuzzy-C-Means Algorithm

[49] http://docwiki.embarcadero.com/RADStudio/Rio/en/Help_on_Help#Developer_Support_Services_an d_Web_Site

[50] http://docwiki.embarcadero.com/RADStudio/Rio/en/Finding_information_(IDE_Tutorial)#Web- based_Product_Information

[51] ttps://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc

[52] DICOM Library https://www.dicomlibrary.com/dicom/

[53] TCIA(The Cancer Imaging Archieve) https://www.cancerimagingarchive.net/

[54] BRATS(Brain Tumor Segmentation) http://braintumorsegmentation.org/

[55] kaggle https://www.kaggle.com/

[56] https://www.slideshare.net/ASHI14march/image-pre-processing

[57] https://www.cs.auckland.ac.nz/courses/compsci373s1c/PatricesLectures/Image%20Filtering_2up.pdf

[58] https://pdfs.semanticscholar.org/9b10/febd207e92f9d362ed06286d0ba31e290fec.pdf

[59] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4879034/

[60] https://www.sciencedirect.com/topics/computer-science/image-segmentation

[61] https://www.sciencedirect.com/topics/computer-science/image-classification

[62] Recommended Practice for Software Design Descriptions,2012, https://sovannarith.files.wordpress.com/2012/07/sdd_template.pdf

[63] IEEE Standard for Information Technology—Systems Design— Software Design Descriptions,2009, http://cengproject.cankaya.edu.tr/wp-content/uploads/sites/10/2017/12/SDD-ieee-1016-2009.pdf

[64] https://github.com/CankayaUniversity/ceng-407-408-2019-2020-Image-Analysis-for-the-Classification-of-Brain-Tumor-Location-on-MR-Images